·论著·

基于人工智能的胸腰椎骨密度测定系统及其校准研究

熊鑫[1,2]，李洋[4]，石峰[4]，杨连[1,2]，段维[1,2]，陈蓓[1,2]，李勇[2]，赵林伟[2]，付泉水[2]，范小萍[2]，杨国庆[3*]

1.637000 四川省南充市，川北医学院医学影像学院
2.629000 四川省遂宁市中心医院放射影像科
3.629000 四川省遂宁市中医院
4.200000 上海市，上海联影智能医疗科技有限公司
*通信作者：杨国庆，主任医师；E-mail：snygq@163.com

【摘要】 **背景** 近年来我国老龄化趋势逐渐增高，骨质疏松的发病率也逐渐攀升，成为了困扰老年人身心健康的主要疾病，而且，国内外诊断及治疗骨质疏松的成本较高，因此骨质疏松的早期诊断成为了降低患者疾病痛苦及治疗成本的关键。**目的** 基于常规胸腹部CT平扫图像，通过深度神经网络和机器学习算法建立胸部和腹部骨密度测定模型，并通过腹部测定模型对胸部骨密度测定结果进行校准，实现自动化的骨密度测量和骨质疏松诊断。**方法** 回顾性收集四川省遂宁市中心医院2022年3月—2023年6月前后一年既行胸部CT检查又完成QCT检查的702例患者为研究对象，其中532例按照随机分组的方式分为训练集（426例，80%）和验证集（106例，20%）。另外170例作为模型的内部测试集。本文以定量CT（QCT）的诊断结果作为参考标准，使用逻辑回归、随机梯度下降及随机森林等机器学习方法构建胸部和腹部的骨质疏松分类模型和骨密度回归模型，同时对建立的模型进行了内部测试，并采用灵敏度、特异度、准确率、精确率及受试者工作特征曲线下面积（AUC）等指标评估模型分类性能，采用平均绝对误差、均方根误差及决定系数等指标评估模型回归性能。**结果** 胸部和腹部的骨质疏松分类模型验证集AUC值分别为0.948和0.968，骨密度回归模型平均绝对误差分别为10.534和9.449；在内部测试集中分类模型AUC值分别为0.905和0.926，回归模型平均绝对误差分别为9.255和7.924；校准后的胸部骨密度测定模型验证集AUC和平均绝对误差分别提高至0.967和10.511。**结论** 基于人工智能的胸部和腰部骨密度测定结果与QCT测定的骨密度具有高度相关性及一致性，可有效诊断骨质疏松症。经校准后的胸部骨密度测定模型也进一步提高了模型在诊断中的性能，为胸部CT平扫在骨质疏松症的机会性筛查中的应用发展提供了巨大潜力。

【关键词】 骨质疏松；骨密度；CT平扫；深度学习；机器学习
【中图分类号】 R 681 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2023.0919

Research on the Measurement System and Calibration of Thoracolumbar Vertebral Density Based on Artificial Intelligence

XIONG Xin[1,2]，*LI Yang*[4]，*SHI Feng*[4]，*YANG Lian*[1,2]，*DUAN Wei*[1,2]，*CHEN Bei*[1,2]，*LI Yong*[2]，*ZHAO LinWei*[2]，*FU Quanshui*[2]，*FAN Xiaoping*[2]，*YANG Guoqing*[3*]

1.College of Medical Imaging，North Sichuan Medical University，Nanchong 637000，China
2.Radiology Department of Suining Central Hospital，Suining 629000，China
3.Suining Traditional Chinese Medicine Hospital，Suining 629000，China
4.Department of Research and Development，Shanghai United Imaging Intelligence Co.，Ltd. Shanghai，200232，China
Corresponding author：YANG Guoqing，Chief Physician；E-mail：snygq@163.com

引用本文：熊鑫，杨国庆，李洋，等. 基于人工智能的胸腰椎骨密度测定系统及其校准研究［J］. 中国全科医学，2025. DOI：10.12114/j.issn.1007-9572.2023.0919.［Epub ahead of print］［www.chinagp.net］

XIONG X，YANG G Q，LI Y，et al. Research on the measurement system and calibration of thoracolumbar vertebral density based on artificial lntelligence［J］. Chinese General Practice，2025.［Epub ahead of print］

【Abstract】 **Background** As China's aging population continues to grow, the incidence of osteoporosis has been steadily increasing, posing a significant health challenge for the elderly population. Furthermore, the high cost of diagnosing and treating osteoporosis highlights the importance of early diagnosis as a key strategy to reduce both patient suffering and healthcare expenses. **Objective** The objective of this study is to develop a chest and abdominal bone mineral density (BMD) measurement model using conventional chest and abdominal CT scans, with deep neural networks and machine learning algorithms. The abdominal BMD model is subsequently employed to calibrate the chest BMD measurements, with the goal of enabling automated BMD measurement and the diagnosis of osteoporosis. Method This retrospective study collected 702 patients from Suining Central Hospital in Sichuan Province who underwent both chest CT scans and QCT examinations during the period from March 2022 to June 2023 (spanning approximately one year) as research subjects.. Among them, 532 patients were randomly divided into a training set (426 cases, 80%) and a validation set (106 cases, 20%). An additional 170 patients were included in the internal testing set. This study uses the diagnostic results of quantitative CT (QCT) as the reference standard and employs machine learning methods such as logistic regression, stochastic gradient descent, and random forest to construct osteoporosis classification models and bone density regression models for the chest and abdomen, the established model was also tested internally. The performance of the classification models was evaluated using sensitivity, specificity, accuracy, precision, and area under the receiver operating characteristic curve (AUC), while regression model performance was assessed using mean absolute error (MAE), root mean square error (RMSE), and R-squared. **Results** The results showed that the AUC values for the osteoporosis classification models in the validation set were 0.948 for the chest model and 0.968 for the abdominal model. The mean absolute errors of the BMD regression models were 10.534 and 9.449, respectively. In the internal testing set, the AUC values for the classification models were 0.905 and 0.926, and the MAE for the regression models were 9.255 and 7.924, respectively. After calibration, the AUC and MAE of the chest BMD measurement model in the validation set improved to 0.967 and 10.511, respectively. **Conclusion** The AI-based chest and abdominal BMD measurements demonstrate a high correlation and consistency with QCT measurements, effectively diagnosing osteoporosis. The calibrated chest BMD measurement model further enhances diagnostic performance and offers significant potential for the application of chest CT scans in opportunistic osteoporosis screening.

【Key words】 Osteoporosis; Bone mineral density; CT plain scan; Deep learning; Machine learning

骨质疏松是一种慢性骨骼疾病，其特点是低骨密度（Bone Mineral Density，BMD）和骨组织微结构恶化[1]。随着我国老龄化趋势的逐渐增高，骨质疏松的发病率逐渐攀升，且国内外诊断及治疗骨质疏松的成本较为高昂，甚至有超过一半的骨质疏松患者从未接受过骨质疏松筛查[2]。因此骨质疏松的早期诊断成为了降低患者疾病痛苦及治疗成本的关键。目前，骨质疏松的早期诊断主要依靠骨密度的测定，常用的骨密度测量方法包括定量超声系统（Quantitative Ultrasound System，QUS）、定量计算机断层扫描（Quantitative Computed Tomography，QCT）和双能X射线骨密度仪（Dual energy X-ray Absorptiometry，DXA）[3]。但由于QUS存在检查准确性能低，QCT需要模具及复杂的后处理，DXA忽略了骨骼的形态和结构并且人均设备量不足等缺点，国内外许多学者提出通过人工智能在CT平扫图像上测量骨密度的方法克服上述检查手段存在的局限性，但建立出来的模型通常只进行了简单的验证校准，模型诊断性能无法得到保证。根据《2018版中国定量CT（QCT）骨质疏松症诊断指南》[4]，由于QCT测量的是真正的体积骨密度，能够更敏感地反应骨质疏松的骨密度变化，而且不受脊柱增生退变和血管钙化等因素的影响，因此本研究以QCT的测量结果作为“金标准”，基于CT平扫图像通过人工智能技术分别建立胸部和腹部骨密度测定模型，并通过腹部测定模型对胸部的骨密度测定进行校准，期望能为提升模型诊断性能的同时为胸部CT平扫在骨质疏松症的机会性筛查中的应用发展提供潜力，更好促进人工智能技术在骨质疏松诊断方面的发展，并为其他疾病的诊断模型建立提供新的思路。

1 材料与方法

1.1 实验材料

1.1.1 实验对象：回顾性收集四川省遂宁市中心医院2022年3月—2023年6月前后一年既行胸部CT检查又完成QCT检查的患者。排除标准：（1）既往有脊柱外科手术史；（2）存在椎体骨折；（3）骨岛、骨囊肿或肿瘤等造成椎体局限性骨质密度增高或减低；（4）严重脊柱侧弯等脊柱畸形的患者；（5）运动、呼吸及金属伪影等干扰造成CT图像质量不佳。本研究已通过遂宁市中心医院伦理委员会批准（伦理审批号：KYLLKS20230143）。

1.1.2 实验设备及其相关参数：胸部CT扫描参数为120 KV，60 mA（自动调整），扫描层厚为0.6 mm，扫

描范围从肺尖至肺底。QCT 扫描参数为 120 KV，35 mA（自动调整），扫描层厚为 0.6 mm。腰椎 CT 数据经 CT 工作站处理后传至 QCT-3000 bone densitometry(Image Analysis，Inc) 分析软件计算出骨密度值（BMDQCT）。

1.2 实验方法

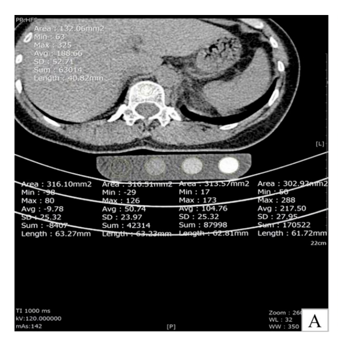
1.2.1 CT 图像采集及 QCT 图像后处理：根据排除标准，在影像存档与通信系统（Picture Archiving and Communication System，PACS）系统中筛选检索到相关患者并将符合要求的患者胸部 CT 平扫图像和 QCT 检查原始平扫图像以 DICOM 格式完整拷贝并导出，以患者姓名及扫描部位为文件夹名进行分类汇总。所有原始的 QCT 平扫图像均由 CT 后处理工作站进行重建后，依次选取腰 1（L1）~ 腰 3（L3）椎体中心层，用尽可能大的感兴趣区域（Region of Interest，ROI）测量患者下方含有标准密度材料体模的平均 CT 值，然后用同样大小的 ROI 测量同层面椎体骨松质的平均 CT 值，椎体 ROI 避开骨皮质及椎基静脉孔，最后由骨密度分析软件 QCT-3000 Densitometry 2.0（Image Analysis，Inc.）计算得到每块椎体的骨密度及平均骨密度值。如遇骨折或存在病变的椎体，则选择相近的椎体进行测量；质量控制分析采用校准体模进行，该校准模型由模拟椎体成分的等效固体材料组成，3 个标准椎体的骨松质的羟基磷灰石密度分别为 50 mg/cm3、100 mg/cm3 和 200 mg/cm3。CT 图像采集及 QCT 图像后处理如下图 1 所示。

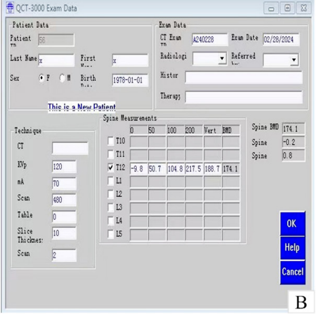
1.2.2 数据收集：收集相关病例影像数据的同时记录患者检查号、性别、年龄以及 QCT 测量的 BMD 值。根据《2018 版中国定量 CT 骨质疏松症诊断指南》[4]进行诊断，如果 BMD 值小于 80 mg/cm^3 诊断为骨质疏松，80~120 mg/cm^3 为骨量降低，超过 120 mg/cm^3 为骨量正常。

1.2.3 脊柱分割及提取：纳入信息脱敏后的 CT 平扫影像序列，通过上海联影智能医疗科技有限公司开发的科研平台系统（uAI Research Portal，uRP）进行脊柱分割及特征提取，uRP 是一款为科学研究而设计的医学图像综合分析软件，支持医学图像可视化及自动分割、配准等高级分析技术。其中 uRP 上内置的分割算法集成了 VB-Net 网络预训练的分割模型，V-Net 通过使用残差函数和跳跃连接，使其在分割精度方面优于各种传统的卷积神经网络模型[5]，而 VB-Net 基于 V-Net 架构通过引入瓶颈结构进一步增强了其分割能力[6]。本研究基于 uRP 椎体分割模型得到椎体的自动分割结果，再对椎体自动分割结果进行腐蚀 3 mm 处理以获取椎体的松质骨，并将其作为 ROI 来提取后续建模所需特征。腹部 CT 平扫矢状面图像及椎体分割结果如图 2 所示。

1.2.4 骨密度测定模型构建：骨密度测定模型包括骨质疏松分类模型和骨密度回归模型，将 QCT 骨密度测量结果作为“金标准”，模型训练阶段将每块椎体松质骨 ROI 的 CT 值均值作为椎体特征，并纳入性别和年龄信息作为临床特征，得到胸部和腹部的骨质疏松分类和骨密度回归模型。包含的主要过程为，计算分割后不同椎体 ROI 的平均 *HU* 值，将性别、年龄和 ROI HU 均值作为特征，QCT 测量的骨密度分类结果作为标签分别输入 Bagging 决策树（Bagging Decision Tree，BDT）、Logistic 回归（Logistic Regression，LR）和随机森林（Random Forest，RF）三种算法建立骨质疏松的分类模型；将 QCT 测量的骨密度值作为标签分别输入 BDT、随机梯度下降（Stochastic Gradient Descent，SGD）和 RF 三种算法建立骨密度回归模型。特征输入模型之前，均采用 Z 分数标准化方法，将不同的数据转换到相同的量级上，基于上述步骤构建胸部模型（T5-T10）和腹部模型（T12-L2）。

1.2.5 模型校准：由于“金标准”骨密度值是通过 QCT 测量 L1 至 L3 或相邻椎体获得，而胸部模型（T5-T10）只包含胸椎椎体特征，缺少腰椎的特征信息。因此，本研究基于胸腰椎不同椎体之间的 *HU* 值差异，对 T5 至 L2 的每块椎体的平均 *HU* 值进行回归拟合，从而获得胸腰椎体 *HU* 值的校准曲线，通过该校准曲线对胸部骨

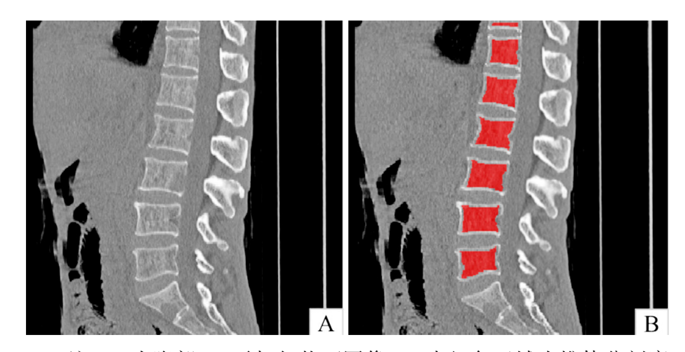




注：A 为定量计算机断层扫描（QCT）测量中校准体模和 CT 平扫图像，B 为 QCT 软件的骨密度值计算界面。

图 1 QCT 骨密度测量图

Figure 1 QCT bone density measurement images



注：A 为腹部 CT 平扫矢状面图像，B 中红色区域为椎体分割腐蚀后的松质骨。

图 2 CT 平扫椎体分割图像

Figure 2 Vertebral segmentation image of CT plain scan

密度测定模型进行校准。具体校准方法如下，首先计算T5至L2每块椎体的平均 *HU* 值，利用线性回归方法得到椎体间 *HU* 值的平均变化系数k，通过系数k和截距项Intercept计算校准后的胸椎椎体 *HU* 值 HU_{cal}，校准公式如下：

$$HU_{cal}=k*\mathrm{vertebra}_n+\mathrm{Intercrpt} \quad 式(1)$$

其中 *k* 是椎体 *HU* 值的变化系数，$\mathrm{vertebra}_n$ 是椎体序号（T5–L2：1–10），Intercrpt是截距项，HU_{cal} 是校正后的椎体 *HU* 值。通过以上校准方法，将计算得到的腰椎特征参数应用到胸部模型中，对胸部骨密度测定模型进行校准。整体实验方法流程图如下图3所示。

1.3 统计学方法

本研究中模型的构建和统计分析采用Python 3.7.4和R 4.1.2软件。符合正态分布的计量资料以（$\bar{x}\pm s$）表示，2组间比较采用独立样本 *t* 检验，非正态性分布的计量资料采用 M（P_{25}，P_{75}）表示，3组间比较采用Kruskal-Wallis *H* 检验。计数资料用相对数表示，组间比较采用 χ^2 检验。绘制受试者工作特征曲线（Receiver Operating Characteristic curve，ROC），计算ROC曲线下面积（Area Under Curve，AUC）、灵敏度、特异度、准确率、精确率和F1值来评估骨质疏松分类模型的性能。通过平均绝对误差（Mean Absolute Error，MAE）、均方根误差（Root Mean Squared Error，RMSE）、可释方差得分（Explained Variance Score，EVS）及决定系数（R-Squared）来评估骨密度回归模型的性能，并通过Pearson相关分析和Bland-Altman探讨模型测定的BMD与金标准之间相关性。以上所有统计检验均为双侧检验，*P*<0.05认为差异有统计学意义。

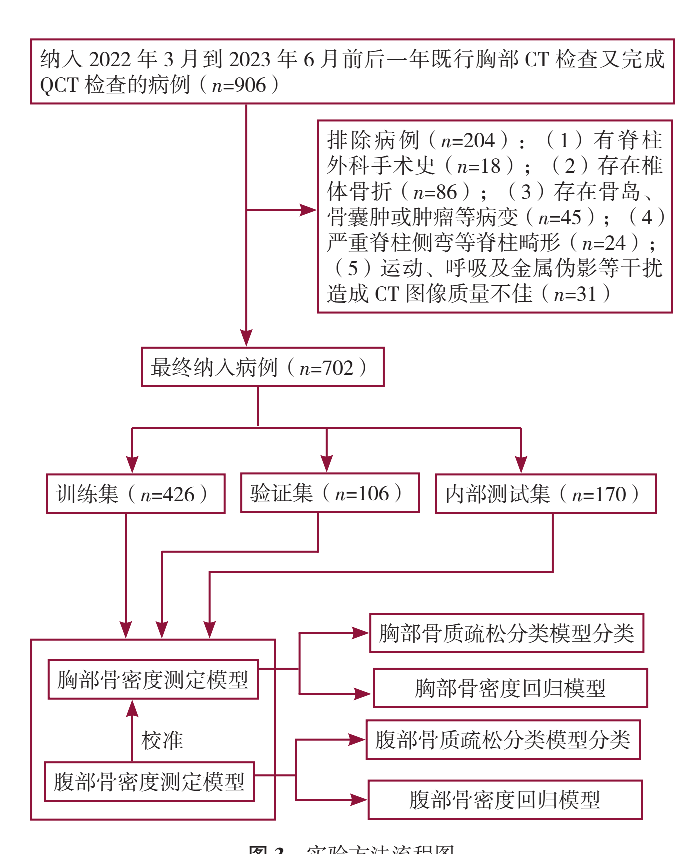


图3 实验方法流程图

Figure 3 Experimental method flow chart

2 结果

2.1 人口统计资料情况

从PACS系统中共收集702例患者的CT平扫图像，其中2022年3月到2023年2月的532例按照随机分组的方式分为训练集426例（80%）和验证集106例（20%）。2023年3月到2023年6月的170例作为模型的内部测试集。训练集、验证集与内部测试集的年龄和性别比较，差异有统计学意义（*P*<0.05），训练集、验证集与内部测试集的骨密度比较，差异无统计学意义（*P*>0.05），见表1。

2.2 骨密度测定模型的建立及评估

2.2.1 骨质疏松分类模型：BDT模型效果最佳，在胸部（T5–T10）和腹部（T12–L2）中验证集AUC值分别为0.948和0.968，胸部分类模型经校准后（T5–T10–Cal）AUC值达到0.967，各分类模型的具体性能指标结果如表2所示，BDT在验证集中的ROC曲线图如图4所示。

2.2.2 骨密度回归模型：SGD模型效果最佳，在胸部（T5–T10）和腹部（T12–L2）中验证集MAE分别为10.534和9.449，胸部回归模型经校准后（T5–T10–Cal）MAE值降低至10.511，各回归模型的具体性能指标结果如表3所示，以上差异均有统计学意义（*P*<0.05）。Pearson相关性分析和Bland-Altman图也显示出SGD模型与QCT测量的BMD之间具有良好的一致性。SGD在验证集中的相关性曲线图如下图5所示，Bland-Altman分析图如下图6所示。

2.3 骨密度测定模型的内部测试

170例内部测试集在T5–T10、T12–L2和T5–T10–Cal的BDT骨质疏松分类模型中，AUC值分别为0.905、0.926和0.918，具体性能指标结果如表4所示，ROC曲线图如图7所示；回归模型MAE分别为9.255和7.924，各模型的具体性能指标结果如表5所示，以上差异均具有统计学意义（*P*<0.05），同时根据Pearson相关性分析显示表明，所开发的三种回归模型与胸腹部测定模型之间具有良好的相关性，如图8所示。

3 讨论

既往关于基于人工智能建立胸部和腹部骨密度测定并进行校准的研究并不多，国内外大多数研究是对已建立的诊断模型进行了简单的验证。PAN等[7]计算每块椎体水平（T12–L2）圆柱体感兴趣体积（Volume of

表 1　训练集、验证集及内部测试集人口统计资料情况

Table 1　Demographic tables for training，validation，and internal testing sets

项目	训练集（*n*=426）	验证集（*n*=106）	内部测试集（*n*=170）	χ^2（H）值	*P* 值
年龄［M（P_{25}，P_{75}），岁］	66（57，73）	65（55，72）	59（55，69）	18.424[a]	<0.001
性别［例（%）］				14.417[b]	0.001
男	160（37.6）	31（29.2）	37（21.8）		
女	266（62.4）	75（70.8）	133（78.2）		
骨密度［例（%）］				2.614b	0.624
骨质疏松	155（36.4）	38（35.8）	56（32.9）		
骨量减低	178（41.8）	45（42.5）	83（48.8）		
正常	93（21.8）	23（21.7）	31（18.2）		

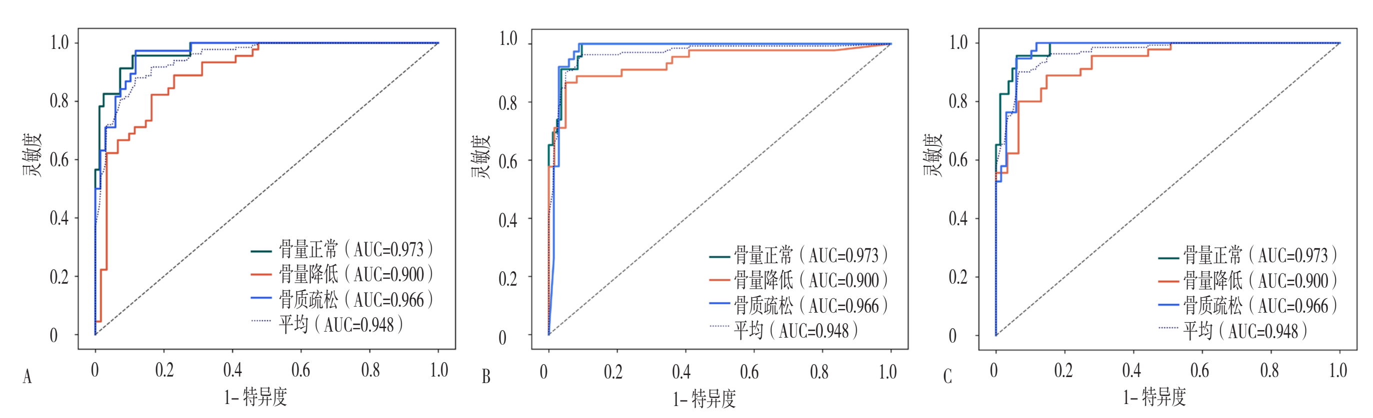
注：[a] 表示 H 值，[b] 表示 χ^2 值。

表 2　BDT、LR 和 RF 建立的骨质疏松分类模型在训练集和验证集中的各性能指标结果

Table 2　Performance index results of osteoporosis classification model established by BDT，LR and RF in training set and validation set

项目	数据集	方法	AUC（95%CI）	灵敏度	特异度	准确率	精确率	F1 值
胸部（T5–T10）	训练集	BDT	0.948（0.929~0.966）	0.793	0.894	0.805	0.824	0.805
	验证集		0.948（0.908~0.984）	0.800	0.900	0.821	0.856	0.819
	训练集	LR	0.891（0.860~0.922）	0.734	0.860	0.735	0.737	0.734
	验证集		0.947（0.909~0.983）	0.801	0.900	0.811	0.817	0.803
	训练集	RF	0.897（0.866~0.927）	0.760	0.875	0.765	0.774	0.766
	验证集		0.942（0.896~0.985）	0.830	0.909	0.830	0.839	0.834
腹部（T12–L2）	训练集	BDT	0.964（0.949~0.978）	0.851	0.920	0.852	0.862	0.856
	验证集		0.968（0.935~0.996）	0.899	0.945	0.896	0.896	0.897
	训练集	LR	0.929（0.904~0.954）	0.803	0.894	0.800	0.806	0.804
	验证集		0.961（0.924~0.993）	0.876	0.934	0.877	0.882	0.878
	训练集	RF	0.915（0.887~0.942）	0.759	0.874	0.763	0.770	0.762
	验证集		0.956（0.917~0.992）	0.866	0.933	0.877	0.891	0.875
胸部分类模型经校准后（T5–T10–Cal）	训练集	BDT	0.955（0.937~0.972）	0.820	0.910	0.833	0.842	0.829
	验证集		0.967（0.938~0.993）	0.851	0.928	0.868	0.884	0.863
	训练集	LR	0.948（0.929~0.967）	0.795	0.895	0.808	0.826	0.807
	验证集		0.949（0.910~0.985）	0.786	0.894	0.811	0.850	0.806
	训练集	RF	0.891（0.860~0.922）	0.734	0.860	0.735	0.737	0.734
	验证集		0.947（0.909~0.983）	0.801	0.900	0.811	0.817	0.803

注：BDT=Bagging 决策树，LR=Logistic 回归，RF= 随机森林；AUC= 受试者工作特征曲线下面积。



注：A 为 Bagging 决策树（BDT）在 T5–T10 验证集中受试者工作特征（ROC）曲线，B 为 BDT 在 T12–L2 验证集中 ROC 曲线，C 为 BDT 在 T5–T10–Cal 验证集中 ROC 曲线。

图 4　BDT 在验证集中各模型的 ROC 曲线图

Figure 4　ROC curves of BDT model in the validation set



表 3 BDT、SGD 和 RF 建立的骨密度回归模型在训练集和验证集中的各性能指标结果
Table 3 Performance index results of osteoporosis regression model established by BDT，SGD and RF in training set and validation set

项目	数据集	方法	MAE	RMSE	EVS	R-Squared	ρ 值	*P* 值
胸部（T5-T10）	训练集	BDT	13.405	18.359	0.743	0.743	0.868	<0.001
	验证集		12.513	15.742	0.793	0.783	0.894	<0.001
	训练集	SGD	12.385	17.460	0.767	0.767	0.876	<0.001
	验证集		10.534	13.352	0.851	0.844	0.923	<0.001
	训练集	RF	13.414	18.268	0.745	0.745	0.866	<0.001
	验证集		12.632	15.984	0.788	0.776	0.888	<0.001
腹部（T12-L2）	训练集	BDT	12.463	16.835	0.784	0.784	0.891	<0.001
	验证集		11.985	14.884	0.818	0.806	0.907	<0.001
	训练集	SGD	11.050	15.147	0.825	0.825	0.909	<0.001
	验证集		9.449	11.563	0.888	0.883	0.942	<0.001
	训练集	RF	12.448	16.374	0.795	0.795	0.894	<0.001
	验证集		12.633	15.837	0.796	0.780	0.892	<0.001
胸部回归模型经校准后（T5-T10-Cal）	训练集	BDT	13.573	18.465	0.740	0.740	0.863	<0.001
	验证集		12.812	16.006	0.787	0.775	0.888	<0.001
	训练集	SGD	12.354	17.471	0.767	0.767	0.876	<0.001
	验证集		10.511	13.312	0.852	0.845	0.923	<0.001
	训练集	RF	13.471	18.330	0.743	0.743	0.864	<0.001
	验证集		12.802	16.079	0.785	0.773	0.887	<0.001

注：MAE= 平均绝对误差，RMSE= 均方根误差，EVS= 可释方差得分；R-Squared= 决定系数。

表 4 BDT 骨密度分类模型在内部测试集的各性能指标结果
Table 4 Performance index results of osteoporosis classification model established by BDT in internal testing sets

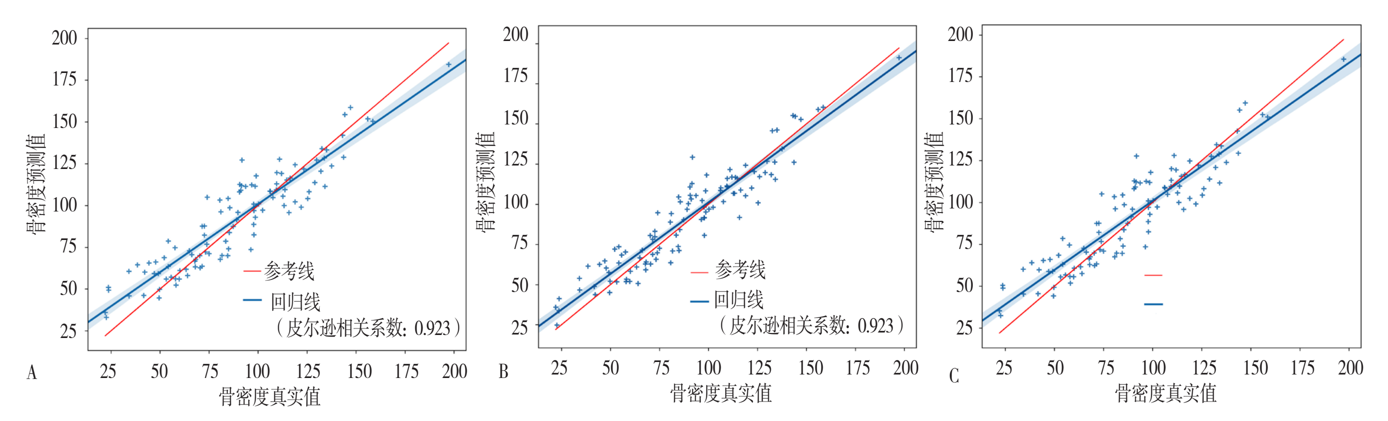
	AUC	灵敏度	特异度	准确率	精确率	F1 值
T5-T10	0.905（0.860-0.947）	0.713	0.869	0.782	0.794	0.734
T12-L2	0.926（0.886-0.965）	0.808	0.918	0.865	0.881	0.829
T5-T10-Cal	0.918（0.878-0.956）	0.746	0.885	0.806	0.814	0.766

表 5 SGD 骨密度回归模型在内部测试集的各性能指标结果
Table 5 Performance index results of osteoporosis regression model established by SGD in internal testing sets

	MAE	RMSE	EVS	R-Squared	ρ	*P* 值
T5-T10	9.255	11.435	0.819	0.818	0.905	<0.001
T12-L2	7.294	9.293	0.883	0.880	0.940	<0.001
T5-T10-Cal	9.248	11.429	0.819	0.818	0.905	<0.001

Interest，VOI）内的平均 CT 值，使用一次线性函数将其映射为 QCT 测量的 BMD 值（mg/cm^3），并通过线性回归方式证明了两者之间具有良好的相关性；JIN 等[8]将各胸椎、第一腰椎 CT 值与股骨颈、全髋、第一至第四腰椎的 t 评分联系起来，通过统计学方式证明了 CT 值与 t 评分呈正相关，且 CT 值与第一腰椎、左髋关节 t 评分相关性较强；LI[9]等通过测量腰椎的 CT 的 HU 平均值，并与 BMD 和 DXA 获得的 t 评分进行比较，生成 ROC 曲线验证模型性能，证明了 CT 平均 HU 与 BMD、T-score 有较好的相关性；ROCK 等[10]通过人工智能算法从胸部 CT 研究中独立定位胸椎，并通过依赖于千伏峰值（kilovolt Peak，KVP）的频谱校正自动计算平均 *HU* 值，表明了该人工智能算法可以成功地确定 BMD 与 DEXA 的适度相关；FANG 等[11]通过使用人工智能对患者椎体的自动分割，以人工勾画的椎体区域作为对照。采用卷积神经网络进行骨密度计算，以 QCT 后处理的值作为分析标准，结果表明了对于四个腰椎（L1-L4）的自动分割和手动分割具有很好的相关性；RÜHLING 等[12]通过分析 C2-T12 各椎体的 vBMD 与腰椎（L1-L3）平均值的关系，并采用线性回归计算骨质疏松症的各椎体特异性临界值，证明了颈胸椎 vBMD 值与腰椎（L1-L3）具有高度的相关性等。虽然国内外研究基于不同方法测量骨密度取得了良好的诊断效果，促进了人工智能技术在医疗行业的发展，但大部分研究仅仅局限于对骨密度诊断模型的建立，并没有对模型进行校准来提高模型的准确性。本研究在建立胸腰椎骨密度测定模型的基础上进一步校准模型来提高模型的准确性。

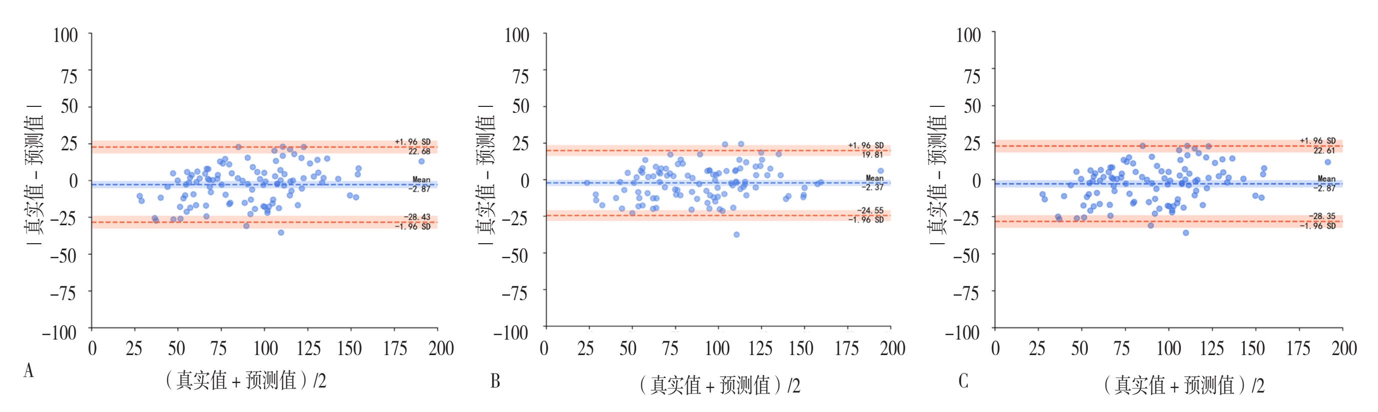
本研究基于人工智能建立胸部和腹部骨密度测定模型并对胸部模型进行校准，实现了自动化骨密度的测量和骨质疏松的诊断。通过对患者相关信息及检查图像进行收集分析、分类及后处理，通过科研平台系统对收集到的图像进行脊柱分割及特征提取，将人工智能技术应用于胸腹部平扫 CT 图像中，通过决策树、逻辑回归、随机梯度下降及随机森林等方法构建了骨质疏松分类模



注：A 为随机梯度下降（SGD）在 T5-T10 验证集中相关性曲线图，B 为 SGD 在 T12-L2 验证集中相关性曲线图，C 为 SGD 在 T5-T10-Cal 验证集中相关性曲线图。

图 5　SGD 在验证集中各模型的相关性曲线图

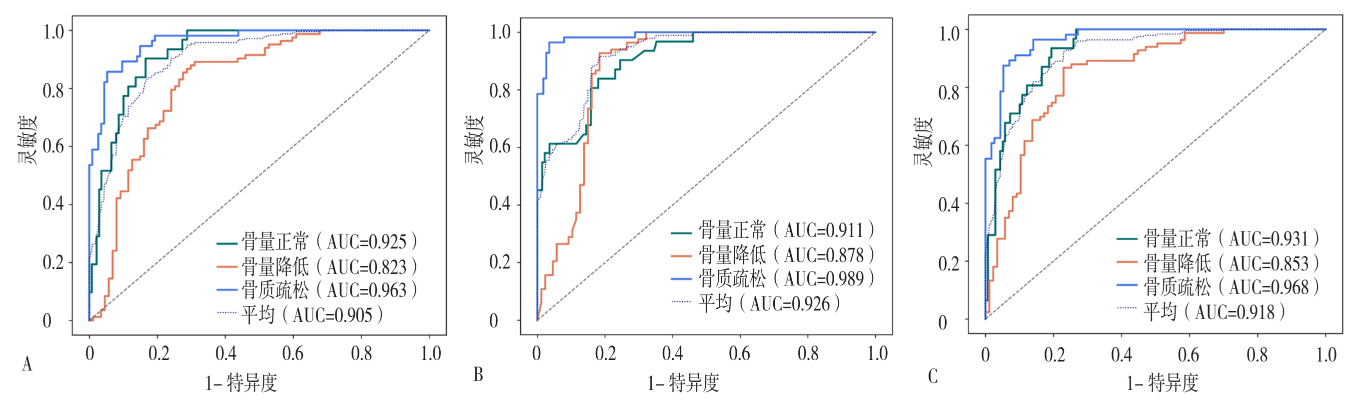
Figure 5　Correlation curves of SGD model in the validation set



注：A 为 SGD 在 T5-T10 验证集中 Bland-Altman 图，B 为 BDT 在 T12-L2 验证集中 Bland-Altman 图，C 为 BDT 在 T5-T10-Cal 验证集中 Bland-Altman 图。

图 6　SGD 在验证集中各模型的 Bland-Altman 图

Figure 6　Bland-Altman diagram of SGD model in the validation set

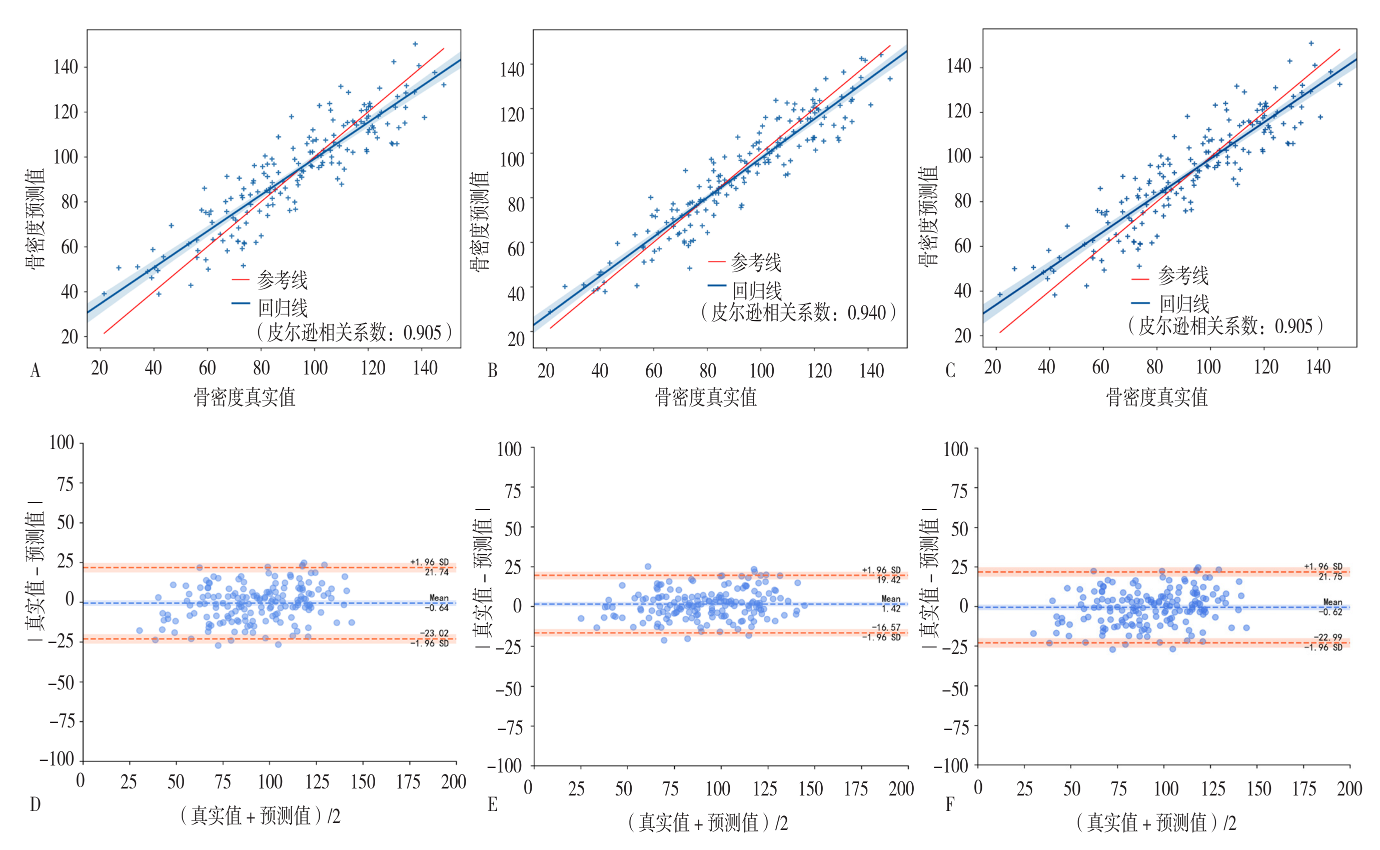


注：A 为 BDT 在 T5-T10 内部测试集中 ROC 曲线，B 为 BDT 在 T12-L2 内部测试集中 ROC 曲线，C 为 BDT 在 T5-T10-Cal 内部测试集中 ROC 曲线。

图 7　BDT 骨密度分类模型在内部测试集中的 ROC 曲线图

Figure 7　ROC curves of BDT model in internal testing sets

型和骨密度回归模型，同时对建立的模型进行了内部测试。研究结果表明，胸部和腹部的骨质疏松分类模型测试集 AUC 分别为 0.948 和 0.968，骨密度回归模型 MAE 分别为 10.534 和 9.449，校准后的胸部骨密度测定模型 AUC 和 MAE 分别提高至 0.967 和 10.511。基于人工智能的胸部和腹部骨密度测定结果与 QCT 测定的骨密度具有高度相关性及一致性，可有效诊断骨质疏松症。经校准后的胸部骨密度测定模型也进一步提高了模型在诊



注：A 为 SGD 在 T5–T10 内部测试集中相关性曲线图，B 为 SGD 在 T12–L2 内部测试集中相关性曲线图，C 为 SGD 在 T5–T10–Cal 内部测试集中相关性曲线图，D 为 SGD 在 T5–T10 内部测试集中 Bland–Altman 图，E 为 SGD 在 T12–L2 内部测试集中 Bland–Altman 图，F 为 SGD 在 T5–T10–Cal 内部测试集中 Bland–Altman 图。

图 8　SGD 骨密度回归模型在内部测试集的相关性曲线图和 Bland–Altman 图

Figure 8　Correlation curves and Bland–Altman diagram of SGD model in internal testing sets

断中的性能，为胸部 CT 平扫在骨质疏松症的机会性筛查中的应用发展提供了巨大的潜力。

本研究使用 QCT 代替传统腰椎的 DXA 作为骨密度的参考标准，为建立的骨密度测定模型性能进行更可靠的评估。虽然腰椎 DXA 是诊断骨质疏松症最常用的参考标准[13]，大多数研究也采用 DXA 来评估 CT 值对诊断骨质疏松症或低骨密度的效能[14–16]，但 DXA 的 BMD 值不是由 CT 值得出的，且 DXA 容易发生腹主动脉钙化和脊柱退变[17]。本研究所建立的骨密度测定模型的 BMD 值与 QCT 获得的 BMD 值具有良好的相关性与一致性，结果表明骨密度测定模型能够实现对椎体骨密度的高精度测量。

本研究通过胸腰椎 *HU* 值校准曲线对胸部骨密度测定模型进行校准。由于胸部模型（T5–T10）只包含胸椎体的特征，缺少腰椎的特征信息，导致胸部模型对骨密度测定的准确率相对较低，通过校准方法，提升了胸部模型（T5–T10）骨密度测定的准确性，这种校准思路，尚未见国内外有过类似文献报道，这为随后的疾病诊断模型校准研究提供了新思路，可促进诊断模型性能的改善及提升。

4　小结

本研究基于机器学习和深度学习方法建立了胸腰椎骨密度测定模型，并使用胸腰椎 *HU* 值校准曲线对已建立的胸部骨密度测定模型进行校准，提高了骨密度测定模型的性能，为人工智能在疾病诊断方向的应用提供了新的思路。随后也将继续已有的研究成果，不断完善人工智能在骨密度测定中的应用，促进人工智能技术在医疗行业的发展。本研究的骨密度测定模型能够满足胸部、腹部及胸腹部 CT 检查患者的检查需求，对相应部位的椎体骨密度进行全自动测量，相较于传统的 QCT 及 DXA 检查，该骨密度测定模型不需要繁琐的操作流程及后处理，只需将检查图像输入到模型集成的软件中，便可得出测量结果，患者无需承受额外的 X 射线剂量，有利于无症状患者的早期筛查，减轻了患者的经济负担；同时，本研究通过模型校准对胸部骨密度测定模型的性能进一步改善和提升，为其它疾病的诊断模型的建立及校准提供了新的思路与方法。

本研究尚存在一些不足之处：（1）本研究是回顾性的，这可能导致选择偏倚，未来可引入 DXA 等检查

结果进行相关研究，未来的前瞻性研究是必要的。（2）本研究的方法是基于单个中心获得的，研究结果及结论还需要进行多中心、更大规模的训练及测试集数据进行研究，以提高研究结果及结论的普适性、稳健性和可重复性。（3）本研究结果及结论是基于 CT 平扫图像作为输入数据建立的，不能应用于增强 CT 图像，未来可以研究增强前后骨密度的变化情况并构建相应模型，甚至可以研究骨密度测定模型在 MRI 中的应用价值。（4）本研究的校准方式是简单的线性拟合，校准方式可能还需要考虑其他情况，以进一步完善校准方式。（5）校准时没有对年龄段进行划分，未来可探讨基于人工智能下不同年龄段的胸腰椎骨密度测定模型及校准情况的不同变化。

作者贡献：熊鑫进行论文的构思与设计，数据的收集与整理，论文撰写；李洋负责论文的修订，统计学处理；李勇等人负责数据的提供；杨连、段维、陈蓓协助数据的收集；杨国庆负责文章的质量控制与审查，对文章整体负责，监督管理。

本文无利益冲突。

参考文献

［1］CHENG X G，YUAN H S，CHENG J L，et al. Chinese expert consensus on the diagnosis of osteoporosis by imaging and bone mineral density［J］. Quant Imaging Med Surg，2020，10（10）：2066–2077. DOI：10.21037/qims–2020–16.

［2］SMITH A D. Screening of bone density at CT：an overlooked opportunity［J］. Radiology，2019，291（2）：368–369. DOI：10.1148/radiol.2019190434.

［3］SCHULTZ K，WOLF J M. Emerging technologies in osteoporosis diagnosis［J］. J Hand Surg Am，2019，44（3）：240–243. DOI：10.1016/j.jhsa.2018.07.006.

［4］《中国定量 CT（QCT）骨质疏松症诊断指南》工作组，程晓光，王亮，等. 中国定量 CT（QCT）骨质疏松症诊断指南（2018）［J］. 中国骨质疏松杂志，2019，25（6）：733–737. DOI：10.3969/j.issn.1006–7108.2019.06.001.

［5］MILLETARI F，NAVAB N，AHMADI S A. V–net：fully convolutional neural networks for volumetric medical image segmentation［C］//2016 Fourth International Conference on 3D Vision（3DV）. October 25–28，2016，Stanford，CA，USA. IEEE，2016：565–571. DOI：10.1109/3DV.2016.79.

［6］HAN M，YAO G，ZHANG W，et al.Segmentation of CT thoracic organs by multi–resolution VB–nets［D］. Proceedings of the 2019 challenge on segmentation of thoracic organs at risk in CT images.

［7］PAN Y L，SHI D J，WANG H Q，et al. Automatic opportunistic osteoporosis screening using low–dose chest computed tomography scans obtained for lung cancer screening［J］. Eur Radiol，2020，30（7）：4107–4116. DOI：10.1007/s00330–020–06679–y.

［8］YANG J R，LIAO M，WANG Y L，et al. Opportunistic osteoporosis screening using chest CT with artificial intelligence［J］. Osteoporos Int，2022，33（12）：2547–2561. DOI：10.1007/s00198–022–06491–y.

［9］LI Y L，WONG K H，LAW M W，et al. Opportunistic screening for osteoporosis in abdominal computed tomography for Chinese population［J］. Arch Osteoporos，2018，13（1）：76. DOI：10.1007/s11657–018–0492–y.

［10］SAVAGE R H，VAN ASSEN M，MARTIN S S，et al. Utilizing artificial intelligence to determine bone mineral density via chest computed tomography［J］. J Thorac Imaging，2020，35（Suppl 1）：S35–S39. DOI：10.1097/RTI.0000000000000484.

［11］FANG Y J，LI W，CHEN X J，et al. Opportunistic osteoporosis screening in multi–detector CT images using deep convolutional neural networks［J］. Eur Radiol，2021，31（4）：1831–1842. DOI：10.1007/s00330–020–07312–8.

［12］RÜHLING S，SCHARR A，SOLLMANN N，et al. Proposed diagnostic volumetric bone mineral density thresholds for osteoporosis and osteopenia at the cervicothoracic spine in correlation to the lumbar spine［J］. Eur Radiol，2022，32（9）：6207–6214. DOI：10.1007/s00330–022–08721–7.

［13］MICHAEL LEWIECKI E，GORDON C M，BAIM S，et al. International society for clinical densitometry 2007 adult and pediatric official positions［J］. Bone，2008，43（6）：1115–1121. DOI：10.1016/j.bone.2008.08.106.

［14］KIM Y W，KIM J H，YOON S H，et al. Vertebral bone attenuation on low–dose chest CT：quantitative volumetric analysis for bone fragility assessment［J］. Osteoporos Int，2017，28（1）：329–338. DOI：10.1007/s00198–016–3724–2.

［15］ROMME E A，MURCHISON J T，PHANG K F，et al. Bone attenuation on routine chest CT correlates with bone mineral density on DXA in patients with COPD［J］. J Bone Miner Res，2012，27（11）：2338–2343. DOI：10.1002/jbmr.1678.

［16］PICKHARDT P J，LEE L J，DEL RIO A M，et al. Simultaneous screening for osteoporosis at CT colonography：bone mineral density assessment using MDCT attenuation techniques compared with the DXA reference standard［J］. J Bone Miner Res，2011，26（9）：2194–2203. DOI：10.1002/jbmr.428.

［17］ENGELKE K，LIBANATI C，LIU Y，et al. Quantitative computed tomography（QCT）of the forearm using general purpose spiral whole–body CT scanners：accuracy，precision and comparison with dual–energy X–ray absorptiometry（DXA）［J］. Bone，2009，45（1）：110–118. DOI：10.1016/j.bone.2009.03.669.

（收稿日期：2024–07–10；修回日期：2025–01–25）

（本文编辑：王世越）